Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5341--98-8301

Threshold Dependent Robust Discrimination for Convex Probability Uncertainty Classes

K.R. Gerlach

Target Characteristics Branch
Radar Division

August 31, 1998

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 31, 1998	

4. TITLE AND SUBTITLE

Threshold Dependent Robust Discrimination for Convex Probability Uncertainty Classes

5. FUNDING NUMBERS

PE-61153N
PR-LR021-05-RH

6. AUTHOR(S)

K.R. Gerlach

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/5341--98-8301

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A methodology for finding robust discriminators for composite binary hypotheses defined for uncertainty classes which are not necessarily 2-alternating capacitable is developed. Past robust discrimination schemes have been threshold independent. In this paper, we present a methodology for finding robust detection structures which are threshold dependent and which sharply upper-bound the Bayes risk over a specified input uncertainty class and the chosen detector threshold. The support of the random variables is assumed to have a finite number of elements. A robust detection structure results from solving an associated limiting minimization problem. Results on the existence of these solutions are presented and conditional solutions for the divergence and divergence/linear uncertainty classes are formulated.

14. SUBJECT TERMS

Detection
Signal Processing
Robust Detection

15. NUMBER OF PAGES

33

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18
298-102

CONTENTS

I. INTRODUCTION 1

II. PRELIMINARIES 2

III. ROBUST SOLUTIONS VIA A MINIMIZATION PROBLEM 4

A. Minimization Problem Definition 4
B. Minimization Solution Convergence 7
C. Minimization Solution Properties 11

IV. EXAMPLE: DIVERGENCE CLASS 14

A. Preliminaries 14
B. Derivation 15
C. Conditions on the Solution 19
D. Calculation of P_D, P_F 21

V. EXAMPLE: DIVERGENCE/LINEAR CLASS 22

VI. SUMMARY 25

REFERENCES 26

APPENDIX A — Proof of Lemma 2 27

APPENDIX B — Proof of Lemma 3 28

APPENDIX C — Condition for Non-Overlapping Divergence Classes 29

APPENDIX D — The divergence or Divergence/Linear Classes Need Not Be 2-alternating Capacitable 30

THRESHOLD DEPENDENT ROBUST DISCRIMINATION FOR CONVEX PROBABILITY UNCERTAINTY CLASSES

I. INTRODUCTION

Robust binary signal discrimination is concerned with finding detection structures whose performance measures over an input class (or classes) are nontrivally lower and/or upper bounded. Normally the underlying probability measures for the binary hypotheses are defined in terms of uncertainty or neighborhood classes. The detector performance measures can be false alarm probability, detection probability, risk, output signal-to-noise (S/N) power ratio, or deflection.

It was proven by Strassen [1, 2] for finite spaces and then by Huber and Strassen [3] for Polish spaces that the Neyman-Pearson lemma generalizes for uncertainty classes that can be characterized as Choquet's 2-alternating capacities [4, 5]. Let Ω be a Polish space and let $\mathscr{F}$ stand for the σ-Borel field on Ω. By M we denote the set of all probability measures on $\mathscr{F}$. The 2-alternating capacity used by Huber and Strassen [3] can be defined as a set function η from $\mathscr{F}$ to [0, 1] which is the upper probability of a weakly compact set of probability measures, and it satisfies the condition $\eta(A \cup B) + \eta(A \cap B) \leq \eta(A) + \eta(B)$ for all $A, B \in \mathscr{F}$. A set $\mathcal{P}$ of all probability measures majorized by η, i.e. $\mathcal{P} = \{P \in M : P(A) \leq \eta(A), \text{ for all } A \in \mathscr{F}\}$, is said to be generated by η.

The Huber-Strassen results were further extended in terms of special capacities by Rieder [6] and Bednarski [7], and general capacities by Vastola [8]. Results for specific uncertainty classes are given by Huber [9, 15], Kassam [10], and Vastola and Poor [11]. All of the above results pertain to the signal discrimination problem whereby each hypothesis is characterized by non-overlapping uncertainty classes. Other approaches consider the generalized signal-to-noise ratio (cf. [12, 13]) as a performance measure.

For all of the results previously cited, the robust test between the two composite hypotheses reduces to a test between two simple hypotheses whereby the underlying probability measures are fixed representatives of the specified uncertainty classes. The representative measures are independent of the test's threshold. In many cases the resultant test is a censored version of a nominal likelihood ratio. Hence, arbitrarily low false alarm probabilities cannot be attained without a trivial randomization of the decision rule. In this paper, we develop a new class of robust discriminators whereby the solutions are threshold dependent (or for short, T-dependent). Specifically we are looking for decision rules such that if the threshold of the rule is specified, then the Bayes risk of the detector is sharply upper-bounded over given input uncertainty classes. By sharp, we mean that there is at least one pair in the hypotheses' uncertainty classes for which the upper-bound is attained. It is in this sense that we define robustness. For this development, the support of the random variables is assumed to have a finite number of elements.

The basic motivation for finding the robust T-dependent solutions is to provide a mechanism for generating robust solutions for uncertainty classes that are not necessarily 2-alternating capacitable. If certain conditions are satisfied, we will find that the uncertainty classes need not be 2-alternating capacitable in order for robust solutions to exist. It will be shown that the robust discrimination solution is again given by a fixed representative pair of simple hypotheses.

This paper is organized as follows. In Section II, we formulate the discrimination problem and summarize an earlier formulation due to Huber [8] and Poor [12]. In Section III, the T-dependent robust solutions for signal discrimination are formulated as solutions of a particular minimization problem similar to that developed by Huber and Strassen [3]. Formulations and conditions for solutions for specific uncertainty classes (the divergence class and its generalization, the divergence/linear class) are given in Sections IV and V.

Manuscript approved August 5, 1998.

II. PRELIMINARIES

Let $(X, \mathscr{F})$ be a measurable space, and let P_0, P_1 be distinct probability measures on it. Assume that P_0 and P_1 are members of two disjoint classes, $\mathcal{P}_0$ and $\mathcal{P}_1$, respectively, of possible distributions on $(X, \mathscr{F})$, and that $\mathcal{P}_0$ and $\mathcal{P}_1$ are convex probability classes, i.e. if $\hat{P}_i$, $\tilde{P}_i \in \mathcal{P}_i$ then $(1 - v)\tilde{P}_i + v\hat{P}_i \in \mathcal{P}_i$ for $0 \leq v \leq 1$, $i = 0, 1$. Let $P_i(i = 0,1)$ have density p_i with respect to some measure μ and assume that $\mu >> P_0$, P_1 and $P_0 >> P_1$ for all $P_i \in \mathcal{P}_i$ $(i = 0,1)$. For this space, X is the set of possible observations and the support of X has a finite number of elements and is denoted by Ω_x. $\mathscr{F}$ is the σ-algebra of possible observation events. In addition, let $\mathbf{x} = \{x_n, n = 1, 2, \ldots, N\}$ be a sequence of complex $(X \subset \mathbf{C}^N)$ identically distributed (but not necessarily independent) random variables (r.v.'s) defined on $(X, \mathscr{F})$.

On the basis of observing the vector $\mathbf{X} = (X_1, X_2, \ldots, X_N)^T$ where T denotes transpose and X_n is the realization of r.v., x_n. We wish to decide between the following pair of hypotheses concerning $\mathbf{X}$,

$$\begin{aligned} H_0&: \mathbf{X} \sim P_0 \in \mathcal{P}_0 \\ H_1&: \mathbf{X} \sim P_1 \in \mathcal{P}_1 \end{aligned} \tag{2.1}$$

where $\mathbf{X} \sim P$ indicates that the observation vector $\mathbf{X}$ is distributed according to the distribution P.

Let ϕ be any test between $\mathcal{P}_0$ and $\mathcal{P}_1$ accepting $\mathcal{P}_1$ with conditional probability $\phi(\mathbf{X})$ given that $\mathbf{X}$ has been observed. Assume that a cost C_i is incurred only if H_i is falsely rejected $(i = 0, 1)$. The expected costs, or risks are given by

$$\bar{R}(P_0,\phi) = C_0\, E\{\phi \mid H_0 \text{ true}\}, \tag{2.2}$$

$$= C_0 \text{ Prob } \{\phi \text{ accepts } H_1 \mid H_0 \text{ true}\},$$

$$\bar{R}(P_1,\phi) = C_1\, E\{1 - \phi \mid H_1 \text{ true}\}, \tag{2.3}$$

$$= C_1 \text{ Prob } \{\phi \text{ accepts } H_0 \mid H_1 \text{ true}\},$$

where E and Prob denote expectation and probability, respectively. Consider the following minimax testing problems:

$$\min_{\phi \in D} \max_{P_1 \in \mathcal{P}_1} \bar{R}(P_1, \phi) \text{ subject to } \max_{P_0 \in \mathcal{P}_0} \bar{R}(P_0, \phi) \leq \alpha \tag{2.4}$$

and

$$\min_{\phi \in D} \max_{P_0, P_1 \in \mathcal{P}} [\pi_0 \bar{R}(P_0, \phi) + \pi_1 \bar{R}(P_1, \phi)] \tag{2.5}$$

where π_i = Prob $\{H_i\}$ occurs, $\mathcal{P} = \mathcal{P}_0 \times \mathcal{P}_1$, and D denotes the class of all randomized decision rules. The problems described by (2.4) and (2.5) are the minimax Neyman-Pearson and Bayes hypothesis testing criteria, respectively.

When the measures P_i and their respective densities, p_i, are known, then the optimal decision rule for both of the above problems is given by the likelihood ratio test [14]:

$$\phi(\mathbf{X}) = \begin{cases} 1, & p_1(\mathbf{X})/p_0(\mathbf{X}) > T \\ \gamma, & p_1(\mathbf{X})/p_0(\mathbf{X}) = T \\ 0, & p_1(\mathbf{X})/p_0(\mathbf{X}) < T \end{cases} \tag{2.6}$$

where the randomization parameter γ and threshold T are chosen to achieve the desired risk performance. For the Bayes criteria, $\gamma = 0$.

When P_0, P_1 are not known but are members of the disjoint uncertainty classes $\mathcal{P}_0$, $\mathcal{P}_1$, respectively, then Huber and Strassen [3] have shown that if the composite hypotheses can be described in terms of alternating capacities of order 2, then the minimax problems given by (2.4) and (2.5) are solved by an ordinary test between a fixed representative pair $\tilde{P}_0$, $\tilde{P}_1$ of simple hypotheses where $\tilde{P}_i \in \mathcal{P}_i$ $(i = 0, 1)$. Note for their development the condition, $P_0 >> P_1$, was not used, the support space need not have a finite number of elements, and that $\tilde{P}_0$, $\tilde{P}_1$ are independent of the threshold, T. In certain cases they also showed that the alternating capacity condition is necessary.

In our development, we wish to define a new class of robust detectors which depend upon the threshold, T. If certain conditions are satisfied, we will find that the alternating capacity condition is not necessary in order for robust solutions to exist and that again the robust solutions are given by a fixed representative pair of simple hypotheses. Our performance measure for optimality is Bayes risk.

To this end, we restrict $\hat{\phi}$ to take the following form for Bayes tests for some $(\hat{P}_0, \hat{P}_1) \in \mathcal{P}_0 \times \mathcal{P}_1$:

$$\hat{\phi}(\mathbf{X}) = \begin{cases} 1 & \hat{p}_1(\mathbf{X})/\hat{p}_0(\mathbf{X}) > T \\ 0 & \hat{p}_1(\mathbf{X})/\hat{p}_0(\mathbf{X}) \le T. \end{cases} \tag{2.7}$$

For a given threshold T we define the T-dependent risks:

$$R(P_0, \hat{\phi}, T) = C_0 \text{ Prob } \{\phi \text{ accepts } H_1 \mid H_0 \text{ true, } T\} \tag{2.8}$$

$$R(P_1, \hat{\phi}, T) = C_1 \text{ Prob } \{\phi \text{ accepts } H_0 \mid H_1 \text{ true, } T\}. \tag{2.9}$$

Measures that can be directly associated with the risks are the probabilities of detection (the power of the test) and false alarm (the size of the test) which we denote by P_D and P_F, respectively. These are defined as

$$P_D(P_1, \hat{\phi}, T) = \text{Prob } \{\phi \text{ accepts } H_1 \mid H_1 \text{ true, } T\}, \tag{2.10}$$

$$P_F(P_0, \hat{\phi}, T) = \text{Prob}\ \{\phi \text{ accepts } H_1 \mid H_0 \text{ true}, T\}. \tag{2.11}$$

In addition the probability of a missed detection is defined by

$$P_M(P_1, \hat{\phi}, T) = 1 - P_D(P_1, \hat{\phi}, T). \tag{2.12}$$

Let $\tilde{\phi}$ be the likelihood ratio associated with a given pair $(\tilde{P}_0, \tilde{P}_1) \in \mathcal{P}$. For arbitrary input pair $(P_0, P_1) \in \mathcal{P}$, the Bayes risk is defined as

$$R_B(P_0, P_1, \tilde{\phi}, T) = \pi_0\, C_0\, P_F(P_0, \tilde{\phi}, T) + \pi_1\, C_1\, P_M(P_1, \tilde{\phi}, T) \tag{2.13}$$

where

$$T = \frac{\pi_0 C_0}{\pi_1 C_1}. \tag{2.14}$$

For a given T, we desire to find a $\tilde{\phi}$ associated with the pair $(\tilde{P}_0, \tilde{P}_1) \in \mathcal{P}$ such that the following bounding condition is satisfied:

$$R_B(\tilde{P}_0, \tilde{P}_1, \hat{\phi}, T) \geq R_B(P_0, P_1, \hat{\phi}, T) \tag{2.15}$$

for all $P_0, P_1 \in \mathcal{P}$.

Definition: For a given T, a pair $(\tilde{P}_0, \tilde{P}_1)$ is called least favorable in terms of risk and T-dependence with respect to the hypothesis test (2.1), if (2.15) is satisfied where $\tilde{\phi}$ is associated with $\tilde{P}_0$, $\tilde{P}_1$ and T according to Eq. (2.7). The pair is also called the least favorable T-dependent pair.

From Bayes risk theory, the least favorable T-dependent pair also satisfies the following inequality:

$$R_B(\tilde{P}_0, \tilde{P}_1, \phi, T) \geq R_B(\tilde{P}_0, \tilde{P}_1, \tilde{\phi}, T) \tag{2.16}$$

where ϕ is arbitrary [12]. The inequalities given by (2.15) and (2.16) indicate that $(\tilde{P}_0, \tilde{P}_1)$ is a saddle point solution of (2.5).

III. ROBUST SOLUTIONS VIA A MINIMIZATION PROBLEM

A. Minimization Problem Definition

Huber and Strassen [3] showed that the least favorable pair associated with the discrimination problem for composite hypotheses that can be described as 2-alternating capacities can be characterized as the solution of an integral minimization problem. This characterization (or a modification of the form given in [10]) also has been found for pdf banded classes and for other optimization criteria where a minimax solution is desired (cf. [12]). In this section, we show for spaces with a finite number of

elements that the least favorable T-dependent pair can also be characterized as the solution of a minimization problem.

For not necessarily finite support spaces, consider the functional defined by

$$\bar{J}(P_0, P_1) = \int_X F\left[\frac{dP_1}{dP_0}\right] dP_0 \tag{3.1}$$

where P_0, P_1 have been previously defined, dP_1/dP_0 is the Radon-Nikodym derivative of P_1 with respect to P_0, and F is a convex function such that its domain and range are in $\mathbb{R}^+ \bigcup \{0\}$. Set $\mathcal{P} = \mathcal{P}_0 \times \mathcal{P}_1$. Consider the problem of finding

$$\min_{\mathcal{P}} \bar{J}(P_0, P_1). \tag{3.2}$$

Under the assumption that the probability densities of $P_i (i = 0, 1)$ exist and are denoted by p_i, then equation (3.1) can be rewritten as

$$J(p_0, p_1) = \int_X F\left[\frac{p_1}{p_0}\right] p_0 \, d\mu \tag{3.3}$$

where $\bar{J}(P_0, P_1) = J(p_0, p_1)$. Let $\mathcal{P}_i (i = 1, 2)$ donote the sets of probability densities associated with $\mathcal{P}_i$. We see that the minimization problem of (3.2) is equivalent to the problem:

$$\min_{\boldsymbol{p}} J(p_0, p_1) \tag{3.4}$$

where $\boldsymbol{p} = \boldsymbol{p}_0 \times \boldsymbol{p}_1$. We note that $\boldsymbol{p}$ is a bounded subset of space $L^1[\mu] \times L^1[\mu]$ and by using Jensen's inequality, J is bounded from below by $F(1)$. Hence we have a problem of minimizing a convex functional that is bounded from below over a bounded subset of a Banach space. Results related to the existence of this minimum can be found in [16-18] and in particular if $\boldsymbol{p}$ is compact via the Weierstrass theorem. Using a result in Poor [12], we can prove the following existence result if $\boldsymbol{p}$ is compact, but J is not necessarily convex.

Theorem 1: Suppose the class $\mathcal{P}_0 \bigcup \mathcal{P}_1$ is dominated by a σ-finite measure μ on $(X, \mathcal{F})$ and that

$$\frac{d(P_0 + P_1)}{d\mu} > 0 \tag{3.5}$$

almost everywhere (a.e.) $[\mu]$ for all $(P_0, P_1) \in \mathcal{P}$. If $\boldsymbol{p}$ is a compact subset of $L^p[\mu] \times L^p[\mu]$ for some $p \geq 1$ (note $||(p_0, p_1)|| = ||p_0|| + ||p_1||$), then the functional $J(p_0, p_1)$ achieves a minimum on $\boldsymbol{p}$.

Proof: The proof is a slight modification of the proof given by Poor for his Theorem 2 [12]. For his Theorem 2, $F(p_1/p_0) = (p_1/p_0)^2$. If we substitute $F(p_1/p_0)$ for $(p_1/p_0)^2$ in his proof, all the conclusions remain the same and Theorem 1 follows. □

We further restrict F to be monotonically increasing, twice differentiable, greater than or equal to zero, and $F'' \geq 0$ (F'' denotes the second derivative of F). Further restrictions will be placed on F as our development proceeds.

We set

$$p_{0v} = (1 - v)\hat{p}_0 + vp_0, \tag{3.6}$$

$$p_{1v} = (1 - v)\hat{p}_1 + vp_1, \tag{3.7}$$

where $0 \leq v \leq 1$; p_0, $\hat{p}_0 \in \boldsymbol{p}_0$; p_1, $\hat{p}_1 \in \boldsymbol{p}_1$; and note that due to convexity $(p_{0v}, p_{1v}) \in \boldsymbol{p}$. Because the support space of $\mathbf{X}$ has a finite number of elements, henceforth we represent all integrations over Ω_x or subsets of Ω_x as summations. Define the following scalar functions of v on [0, 1]

$$H_0(v) = \sum_{\Omega_x} F\left[\frac{\hat{p}_1}{p_{0v}}\right] p_{0v}, \tag{3.8}$$

$$H_1(v) = \sum_{\Omega_x} F\left[\frac{p_{1v}}{\hat{p}_0}\right] \hat{p}_0, \tag{3.9}$$

$$H_2(v) = \sum_{\Omega_x} F\left[\frac{p_{1v}}{p_{0v}}\right] p_{0v}. \tag{3.10}$$

Lemma 1: H_0, H_1, H_2 are convex functions of v on [0, 1].

Proof: It is straightforward to show that

$$\frac{dH_0}{dv} = \sum_{\Omega_x} (p_0 - \hat{p}_0)\left[F\left[\frac{\hat{p}_1}{p_{ov}}\right] - \frac{\hat{p}_1}{p_{0v}} F'\left[\frac{\hat{p}_1}{p_{0v}}\right]\right]. \tag{3.11}$$

$$\frac{d^2H_0}{dv^2} = \sum_{\Omega_x} \frac{\hat{p}_1(p_0 - \hat{p}_0)^2}{p_{0v}^3} F''\left[\frac{\hat{p}_1}{p_{0v}}\right] \geq 0. \tag{3.12}$$

$$\frac{dH_1}{dv} = \sum_{\Omega_x} (p_1 - \hat{p}_1) F'\left[\frac{p_{1v}}{\hat{p}_0}\right]. \tag{3.13}$$

$$\frac{d^2H_1}{dv^2} = \sum_{\Omega_x} \frac{(p_1 - \hat{p}_1)^2}{\hat{p}_0} F''\left[\frac{\hat{p}_{1v}}{\hat{p}_0}\right] \geq 0. \tag{3.14}$$

$$\frac{dH_2}{dv} = \sum_{\Omega_x} \left\{ (p_0 - \hat{p}_0) \left[F\left[\frac{p_{1v}}{p_{0v}}\right] - \frac{p_{1v}}{p_{0v}} F'\left[\frac{p_{1v}}{p_{0v}}\right] \right] \right. \tag{3.15}$$

$$\left. + (p_1 - \hat{p}_1) F'\left[\frac{p_{1v}}{p_{0v}}\right] \right\}$$

and

$$\frac{d^2H_2}{dv^2} = \sum_{\Omega_x} \frac{(\hat{p}_1 p_0 - p_1 \hat{p}_0)^2}{p_{0v}^3} F''\left[\frac{p_{1v}}{p_{0v}}\right] \geq 0. \tag{3.16}$$

where F' is the first derivative of F. The results of (3.12), (3.14), and (3.16) verify the lemma.□

Because H_2 is a convex function of v it follows that J is convex on $\mathbf{p}$.

B. Minimization Solution Convergence

A useful functional form for F is now defined which allows us to obtain our results. Define the function G on $\mathbb{R}^+ \bigcup \{0\}$ as

$$G(z) = zF'(z) - F(z). \tag{3.17}$$

It can be shown that

$$F(z) = z \int_0^z \frac{G(\beta)}{\beta^2} \, d\beta \tag{3.18}$$

and

$$G'(z) = zF''(z). \tag{3.19}$$

Set $G = G_\epsilon(z, T)$ and define the function

$$G_\epsilon(z,T) = \begin{cases} 0 & \text{for } z \in \Omega_0 \equiv [0,\ T) \\ \dfrac{1}{\epsilon}\,(z - T) & \text{for } z \in \Omega_1 \equiv [T,\ T + \epsilon) \\ 1 & \text{for } z \in \Omega_2 \equiv [T + \epsilon,\ \infty] \end{cases} \tag{3.20}$$

where the sets Ω_0, Ω_1, and Ω_2 are defined in the equation. For this characterization, $G_\epsilon(z,T) \to u(z - T)$ as $\epsilon \downarrow 0$ where u is the Heavyside step function. For this $G_\epsilon\ (z,T)$, we can write $F_\epsilon(z,T)$ using (3.18) explicitly as

$$F_\epsilon(z,T) = \begin{cases} 0 & ,\ z \in \Omega_0 \\ \dfrac{1}{\epsilon}\left[z \ln \dfrac{z}{T} - (z - T)\right] & ,\ z \in \Omega_1 \\ \dfrac{z}{\epsilon} \ln \left[1 + \dfrac{\epsilon}{T}\right] - 1 & ,\ z \in \Omega_2. \end{cases} \tag{3.21}$$

Using (3.21), it is straightforward to show that F_ϵ is twice differentiable with respect to z, greater than or equal to zero, and monotonically increasing. However F_ϵ'' is not continuous. Define

$$F_0(z,T) = \begin{cases} 0 & ,\ z < T \\ \dfrac{1}{T}\,(z - T) & ,\ z \geq T. \end{cases} \tag{3.22}$$

We will need the following two lemmas for our development:

Lemma 2: $F_\epsilon(z,\ T)$ converges uniformly on $[0,\ z_{max}]$ to $F_0(z,\ T)$ as $\epsilon \downarrow 0$ where $0 < z_{max} < \infty$.

Proof: See Appendix A.

Lemma 3: Set $H_\epsilon(z,T) \equiv F_\epsilon'(z,T) - \dfrac{1}{T}\, G_\epsilon(z,T)$. For the characterization of $G_\epsilon(z,T)$ given by (3.20), $H_\epsilon(z,\ T)$ converges uniformly to 0 for $z \geq 0$ as $\epsilon \downarrow 0$.

Proof: See Appendix B.

Define the functional $J_\epsilon(p_0,\ p_1)$ as

$$J_\epsilon(p_0,\ p_1) = \sum_{\Omega_s} p_0\ F_\epsilon(p_1/p_0, T) \tag{3.23}$$

and set (if they exist)

$$(\tilde{p}_0,\ \tilde{p}_1) = \arg \min_{p_0, p_1 \in p} J_0(p_0,\ p_1) \tag{3.24}$$

$$(\tilde{p}_0^\epsilon,\ \tilde{p}_1^\epsilon) = \arg \min_{p_0, p_1 \in p} J_\epsilon(p_0, p_1). \tag{3.25}$$

We point out that $\tilde{p}_i^\epsilon (i = 0,1)$ can also depend on T. The following theorem establishes conditions under which J_ϵ converges as $\epsilon \downarrow 0$.

Theorem 2: If

C1. $\tilde{p}_i^\epsilon \rightarrow \tilde{p}_i$ uniformly on Ω_x as $\epsilon \downarrow 0,\ i = 0,1$

C2. $\min\limits_{\mathbf{x}} \tilde{p}_0$ exists then

$$\tilde{p}_1^\epsilon / \tilde{p}_0^\epsilon \rightarrow \tilde{p}_1 / \tilde{p}_0 \text{ uniformly on } X \text{ as } \epsilon \downarrow 0, \tag{3.26}$$

$$J_\epsilon(\tilde{p}_0^\epsilon, \tilde{p}_1^\epsilon) \rightarrow J_0(\tilde{p}_0, \tilde{p}_1) = T^{-1} - \beta R_B(\tilde{P}_0,\ \tilde{P}_1\ \tilde{\phi},\ T) \tag{3.27}$$

where $\beta = (\pi_0\ C_0)^{-1}$.

Proof: It is elementary to show that C1 and C2 imply (3.26). Define $e(\epsilon) = \sup\limits_{\mathbf{x}} |\tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon - \tilde{p}_1/\tilde{p}_0|$. Eq. (3.26) implies $e(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ and

$$\tilde{p}_1/\tilde{p}_0 - e(\epsilon) \leq \tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon \leq \tilde{p}_1/\tilde{p}_0 + e(\epsilon). \tag{3.28}$$

Define the following sets

$$S = \{\mathbf{x} \mid \tilde{p}_1/\tilde{p}_0 > T\}$$

$$S_\epsilon = \{\mathbf{x} \mid \tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon > T\}$$

$$S_\epsilon^+ = \{\mathbf{x} \mid \tilde{p}_1/\tilde{p}_0 + e(\epsilon) > T\}$$

$$S_\epsilon^- = \{\mathbf{x} \mid \tilde{p}_1/\tilde{p}_0 - e(\epsilon) > T\}$$

$$A = \{\mathbf{x} \mid \tilde{p}_1/\tilde{p}_0 = T\}$$

$$A_\epsilon = S_\epsilon^+ - A.$$

Because of Lemma 2 and $\mu(\Omega_x) < \infty$, we can write

$$\begin{aligned} J_\epsilon(\tilde{p}_0^\epsilon, \tilde{p}_1^\epsilon) &= \sum_{S_\epsilon} \tilde{p}_0^\epsilon \frac{1}{T} \left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon} - T \right] + O(\epsilon) \\ &= \sum_{S_\epsilon} \left[\frac{1}{T} \tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon \right] + O(\epsilon). \end{aligned} \tag{3.29}$$

Define the following set functions

$$\eta^\epsilon(C) = \sum_C \left[\frac{1}{T} \tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon \right] \tag{3.30a}$$

$$\eta(C) = \sum_C \left[\frac{1}{T} \tilde{p}_1 - \tilde{p}_0 \right] \tag{3.30b}$$

where $C \in \mathscr{F}$. Because $S_\epsilon^- \subseteq S_\epsilon \subseteq S_\epsilon^+$, it follows that

$$\eta^\epsilon(S_\epsilon^-) \leq \eta^\epsilon(S_\epsilon) \leq \eta^\epsilon(S_\epsilon^+). \tag{3.31}$$

Using C1-2 and $\mu(\Omega_x) < \infty$, it is straightforward to show that

$$\eta^\epsilon(S_\epsilon^-) = \eta(S_\epsilon^-) + O(\epsilon) \tag{3.32}$$

$$\eta^\epsilon(S_\epsilon) = \eta(S_\epsilon) + O(\epsilon) \tag{3.33}$$

$$\eta^\epsilon(S_\epsilon^+) = \eta(S_\epsilon^+) + O(\epsilon). \tag{3.34}$$

Also with little difficulty, we can show

$$\lim_{\epsilon \downarrow 0} \eta(S_\epsilon^-) = \eta(S) \tag{3.35a}$$

and

$$\lim_{\epsilon \downarrow 0} \eta(A_\epsilon) = \eta(S). \tag{3.35b}$$

Now because $A_\epsilon \cap A = \Phi$ (the empty set), then

$$\eta(S_\epsilon^+) = \eta(A_\epsilon) + \eta(A). \tag{3.36}$$

However,

$$\eta(A) = \Sigma_{\frac{\tilde{p}_1}{\tilde{p}_0} = T} \left[\frac{1}{T}\tilde{p}_1 - \tilde{p}_0\right] = 0. \tag{3.37}$$

Thus $\eta(S_\epsilon^+) = \eta(A_\epsilon)$ and $\lim_{\epsilon \downarrow 0} \eta(S_\epsilon^+) = \eta(S)$.. Using this result and (3.31)-(3.35), we see that

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \eta^\epsilon(S_\epsilon) &= \eta(S) \\ &= \frac{1}{T} P_D(\tilde{P}_1,\tilde{\phi},T) - P_F(\tilde{P}_0,\tilde{\phi},T) \\ &= \frac{1}{T} - \frac{1}{\pi_0\, C_0} R_B(\tilde{P}_0,\tilde{P}_1,\tilde{\phi},T) \ \square \end{aligned} \tag{3.38}$$

C. Minimization Solution Properties

The densities, $\tilde{p}_0^\epsilon$, $\tilde{p}_1^\epsilon$ are defined by (3.25). Using Lemma 1, for $\hat{p}_0 = \tilde{p}_0^\epsilon$ and $\hat{p}_1 = \tilde{p}_1^\epsilon$

$$\left.\frac{dH_0}{dv}\right|_{v=0} \geq 0. \tag{3.39}$$

Using (3.11), (3.17), and (3.39), it follows that (with $G = G_\epsilon\,(z, T)$)

$$\sum_{\Omega_x} \tilde{p}_0^\epsilon\, G_\epsilon \left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] \geq \sum_{\Omega_x} p_0\, G_\epsilon \left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] \text{ for all } P_0 \in \mathcal{P}_0. \tag{3.40}$$

In similar fashion, since Lemma 1 implies

$$\left.\frac{dH_1}{dv}\right|_{v=0} \geq 0, \tag{3.41}$$

it follows via Lemma 3 that

$$\sum_{\Omega_x} \tilde{p}_1^\epsilon\, G_\epsilon \left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] \leq \sum_{\Omega_x} p_1\, G_\epsilon \left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right]. \tag{3.42}$$

We can combine (3.40) and (3.42) to obtain

$$\sum_{\Omega_\epsilon} \left[\frac{1}{T}\tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon\right] G_\epsilon\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] \le \sum_{\Omega_\epsilon} \left[\frac{1}{T}p_1 - p_0\right] G_\epsilon\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right]. \tag{3.43}$$

Using Lemma 3, we can write

$$\sum_{\Omega_\epsilon} \left[\frac{1}{T}\tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon\right] G_\epsilon\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] = \sum_{S_\epsilon} \left[\frac{1}{T}\tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon\right] + O(\epsilon) \tag{3.44}$$

where S_ϵ was previously defined. Thus under the conditions of Theorem 2,

$$\lim_{\epsilon\downarrow 0} \sum_{\Omega_\epsilon} \left[\frac{1}{T}\tilde{p}_1^\epsilon - \tilde{p}_0^\epsilon\right] G_\epsilon\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}, T\right] = T^{-1} - \beta\, R_B\,(\tilde{P}_0, \tilde{P}_1, \tilde{\phi}, T). \tag{3.45}$$

We now upper bound the right side of (3.43). Define

$$I_\epsilon^{(F)} = \sum_{\Omega_\epsilon} p_0\, G_\epsilon(\tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon, T). \tag{3.46}$$

We know

$$I_\epsilon^{(F)} \ge \sum_{B_{n,\epsilon}} p_0 + O(\epsilon)$$

where

$$B_{n,\epsilon} = \left\{\mathbf{x}\ |\tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon > T + \frac{1}{n}\right\}.$$

and n is a positive integer. Assuming condition C2 of Theorem 2, then $\tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon = \tilde{p}_1/\tilde{p}_0 + O(\epsilon)$. Define the set

$$B_{n,\epsilon}^* = \left\{\mathbf{x}\ |\tilde{p}_1/\tilde{p}_0 > T + \frac{1}{n} + e(\epsilon)\right\}$$

and the set functions

$$\eta^0(D) = \sum_D p_0$$

$$\eta^1(D) = \sum_D p_1$$

where $D \in \mathscr{F}$. Since $B^{+}_{n,\epsilon} \subseteq B_{n,\epsilon}$

$$I^{(F)}_{\epsilon} \geq \eta^0(B^{+}_{n,\epsilon}) + O(\epsilon) \tag{3.47}$$

In similar fashion, if we define

$$I^{(M)}_{\epsilon} = \sum_{\Omega_x} p_1(1 - G_{\epsilon}(\tilde{p}^{\epsilon}_1/\tilde{p}^{\epsilon}_0,\ T)), \tag{3.48}$$

we can show

$$I^{(M)}_{\epsilon} \geq \eta^1(C^{+}_{n,\epsilon}) + O(\epsilon) \tag{3.49}$$

where

$$C^{+}_{n,\epsilon} = \left\{ \mathrm{x} \ |\tilde{p}_1/\tilde{p}_0 \leq T - \frac{1}{n} - e(\epsilon) \right\}.$$

Furthermore, using (3.47) and (3.49), it follows that

$$\sum_{\Omega_x} \left[\frac{1}{T} p_1 - p_0 \right] G_{\epsilon}(\tilde{p}^{\epsilon}_1/\tilde{p}^{\epsilon}_0\ ,T) = \frac{1}{T}(1 - I^{(M)}_{\epsilon}) - I^{(F)}_{\epsilon}$$

$$\leq \frac{1}{T} - \frac{1}{T}\eta^1(C^{+}_{n,\epsilon}) - \eta^0(B^{+}_{n,\epsilon}) + O(\epsilon). \tag{3.50}$$

It is straightforward to show that

$$\lim_{n\to\infty} \lim_{\epsilon \downarrow 0} \eta^1(C^{+}_{n,\epsilon}) = P_M(P_1,\ \tilde{\phi},\ T) \tag{3.51}$$

$$\lim_{n\to\infty} \lim_{\epsilon \downarrow 0} \eta^1(B^{+}_{n,\epsilon}) = P_F(P_0,\ \tilde{\phi},\ T). \tag{3.52}$$

Thus combining (3.51), (3.52) with (3.43) and (3.45), it follows that as $\epsilon \downarrow 0$,

$$R_B(\tilde{P}_0,\tilde{P}_1,\tilde{\phi},T) \geq R_B(P_0,P_1,\tilde{\phi},T). \tag{3.53}$$

We summarize the preceding results in the following theorem:

Theorem 3: Assume the pair $(\tilde{p}^{\epsilon}_0,\ \tilde{p}^{\epsilon}_1)$ exists such that

$$(\tilde{p}_0^\epsilon,\ \tilde{p}_1^\epsilon) = \arg\min_{\mathcal{P}} \sum_{\Omega_x} p_0\, F_\epsilon \left[\frac{p_1}{p_0},\ T\right] \tag{3.54}$$

for all $0 \leq \epsilon \leq \epsilon_{\max}$ and some $\epsilon_{\max} > 0$. Under the conditions of Theorem 2, then $(\tilde{P}_0,\ \tilde{P}_1)$ is a least favorable T-dependent pair.

We see from this result and Theorem 1 that under fairly general conditions, least favorable T-dependent solutions exist (note the conditions given for the classes in Theorem 1 are not necessary). In the next sections, we will find conditions and formulations for solutions to the minimization problem posed by (3.4) when $\epsilon > 0$. The least favorable T-dependent solution will result in the limit as $\epsilon \downarrow 0$.

IV. EXAMPLE: DIVERGENCE CLASS

A. Preliminaries

In this section, by way of example, we present a methodology for finding the least favorable T-dependent pair associated with $\mathcal{P}_i(i = 0, 1)$ defined as divergence classes. The divergence [19] of two densities p, q where $Q >> P$ is defined as

$$D(p,q) = \sum_{\Omega_x} p \ln \frac{p}{q}. \tag{4.1a}$$

We define the two hypotheses classes as

$$\mathcal{P}_i = \{P_i \mid D(p_i,p_i^*) \leq \Delta_i\},\ i = 0,1 \tag{4.1b}$$

where p_i^* are the known nominal densities of $H_i\,(i = 0, 1)$, $P_i^* >> P_i$, and Δ_i are positive real numbers chosen such that $\mathcal{P}_0 \cap \mathcal{P}_1 = \Phi$. Note $P_i(i = 0, 1)$ are convex classes. Conditions for $\mathcal{P}_0 \cap \mathcal{P}_1 = \Phi$ are given in Appendix C for $\Delta_0 = \Delta_1$. The divergence class (or simply div-class) is shown not to necessarily be 2-alternating capacitable in Appendix D.

The restriction of the support Ω_x to have a finite number of elements has two significant benefits. The first is related to the fact that in practice all solutions that are formulated have unknown parameters that are solved for via constraint equations and the digital computer. Hence the support is almost always modelled as a finite set in order to solve for the unknown parameters of the T-dependent least favorable pair solution. The second benefit of assuming a finite support set is that one may use the powerful Kuhn-Tucker convexity theorem for finite dimensions [17] that guarantees that if the Lagrange multiplier equation is solvable then the solution is a global minimum. In addition, if the minimum exists on an open subset of the convex probability uncertainty classes, $\mathcal{P}$, then the Lagrange multiplier equation is necessarily solvable.

For this problem, a support, Ω_x, need not be explicitly defined since none of the constraint equations depend on Ω_x. For notational purposes we define an index set, $I = \{1, 2, \ldots, K\}$ where K is the number of elements of Ω_x and $K \geq 3$. Also define

$$p_i = \{p_i \mid D(p_i,p_i^*) \leq \Delta_i,\ p_i \in p^K\},\ i = 0,1\ , \tag{4.2a}$$

$$p_i^{+} = \{p_i \,|D(p_i,p_i^{*}) = \Delta_i \; p_i \in p^K\},\; i = 0,1 \;, \tag{4.2b}$$

$p = p_0 \times p_1$, $p^{+} = p_0^{+} \times p_1^{+}$, where p^K is the space of measures probability measures on a set of K elements. We assume that $p_i^{+}(i = 0,1)$ has an infinite number of elements which will be true if $K \geq 3$.

We observe that $p_i(i = 0,1)$ is a closed subset of hypercube $\mathop{X}_{k=1}^{K} [0,1]$ defined on $\mathbb{R}^K$. Hence, because, $p_i(i = 0,1)$ is closed and bounded then, p_i and p are compact. Thus Theorem 1 guarantees that a minimum to (3.4) exists. Define $p_I^{+} = \text{int}\ (p^{+})$ (int ~ interior of). In the following, we assume for the least favorable T-dependent solution that $(\tilde{p}_0, \tilde{p}_1) \in p_I^{+}$.. Thus since p_I^{+} is an open subset of p^{+}, the Lagrange multiplier equations are necessarily solvable. In addition, under the assumption $(\tilde{p}_0^{\epsilon}, \tilde{p}_1^{\epsilon})$ existing on p^{+}, then $(\tilde{p}_0^{\epsilon}, \tilde{p}_1^{\epsilon}) \in p_I^{+}$ for all $0 \leq \epsilon \leq \epsilon_{\max}$ and some $\epsilon_{\max} > 0$. Thus the Lagrange multiplier equations are necessarily solvable for $(\tilde{p}_0^{\epsilon}, \tilde{p}_1^{\epsilon})$, $0 \leq \epsilon \leq \epsilon_{\max}$, for some $\epsilon_{\max} > 0$.

In order to show that $(\tilde{p}_0, \tilde{p}_1)$ is a least favorable T-dependent pair, we must show that the conditions C1 and C2 of Theorem 2 are met. After $(\tilde{p}_0, \tilde{p}_1)$ is found via the methodology to be presented, one can check and see if $(\tilde{p}_0, \tilde{p}_1) \in p_1^{+}$ and verify condition C2. Finally, we note with respect to condition C1 that for a finite discrete support space of $\mathbf{x}$, pointwise convergence on Ω_x implies uniform convergence on Ω_x.

B. Derivation

For notational purposes, we write $F_\epsilon(z, T) = F_\epsilon(z)$ and $G_\epsilon(z, T) = G_\epsilon(z)$. Consider the minimization of $J(p_0, p_1)$ defined by (3.23) with $p_i \in p_i$ $(i = 0, 1)$ defined by (4.2a). Besides the divergence inequality constraints expressed by (4.2a), the densities must also satisfy the total mass and semi-positivity constraints. Note, all the constraint functions are convex. In constructing the Lagrangian for this minimization problem, we will at first ignore these last constraints and show that the solution obtained without these constraints satisfies the semi-positivity constraints and by proper normalization can be made to satisfy the total mass constraints.

The Lagrangian for this minimization problem on a finite support is given by

$$L = \sum_I \tilde{p}_0^{\epsilon} F_\epsilon' \left[\frac{\tilde{p}_1^{\epsilon}}{\tilde{p}_0^{\epsilon}}\right] + s_0^{\epsilon} \sum_I \tilde{p}_0^{\epsilon} \ln \frac{\tilde{p}_0^{\epsilon}}{p_0^{*}} + s_1^{\epsilon} \sum_I \tilde{p}_1^{\epsilon} \ln \frac{\tilde{p}_1^{\epsilon}}{p_1^{*}} \tag{4.3}$$

where each $\tilde{p}_0^{\epsilon}$, $\tilde{p}_1^{\epsilon}$, p_0^{*} , p_1^{*} is indexed with respect to the elements of I and $s_i^{\epsilon}(i = 0, 1)$ are Lagrange multipliers. We have superscripted these unknowns to indicate that they are functions of ϵ. We will do this with other unknowns as well. We sum over these indexed elements and denote this by $\sum_I$. The Kuhn-Tucker convexity theorem [17] guarantees a global minimum on convex p if the following equations are solvable on p:

$$\frac{\partial L}{\partial \tilde{p}_0^\epsilon} = G_\epsilon\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}\right] + s_0^\epsilon + s_0^\epsilon \ln \frac{\tilde{p}_0^\epsilon}{p_0^*} = 0 \qquad (4.4a)$$

$$\frac{\partial L}{\partial \tilde{p}_1^\epsilon} = F_\epsilon'\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}\right] + s_1^\epsilon + s_1^\epsilon \ln \frac{\tilde{p}_1^\epsilon}{p_1^*} = 0 \qquad (4.4b)$$

$$D(p_i^\epsilon \,,\, p_i^*) = \Delta_i \,,\; (i = 0,1) \qquad (4.5)$$

where $\infty > s_0^\epsilon,\, s_1^\epsilon \geq 0$. If the minimum of J is an interior point of $\boldsymbol{p}$ then (4.4)-(4.5) are necessary and sufficient conditions. In order to obtain a solution we assume $\lim_{\epsilon \downarrow 0} s_i^\epsilon = s_i \geq 0$, $(i = 0, 1)$ and check this condition afterward. We also assume

C3. All of the parameters of the constraint equations which are functions of ϵ have limits as $\epsilon \downarrow 0$.

Later in our development, we will show conditions under which C3 holds for this particular problem. Set $\Lambda^\epsilon = \tilde{p}_1^\epsilon/\tilde{p}_0^\epsilon$ and $\Lambda^* = p_1^*/p_0^*$, where Λ^ϵ and Λ^* represent the solution pair and nominal's likelihood ratios, respectively. Using (4.4a) and (4.4b), we can show

$$\frac{1}{s_0^\epsilon} G^\epsilon(\Lambda^\epsilon) + \frac{1}{s_1^\epsilon} F_\epsilon'(\Lambda^\epsilon) + \ln \Lambda^\epsilon = \ln \Lambda^* . \qquad (4.6)$$

If Λ^ϵ is known then

$$\tilde{p}_1^\epsilon = p_1^* \exp - \left[1 + \frac{1}{s_1^\epsilon} F_\epsilon'(\Lambda^\epsilon)\right] , \qquad (4.7)$$

$$\tilde{p}_0^\epsilon = p_0^* \exp - \left[1 - \frac{1}{s_0^\epsilon} G_\epsilon(\Lambda^\epsilon)\right] . \qquad (4.8)$$

We observe that $\tilde{p}_i^\epsilon \geq 0$, $(i = 0,1)$, so that the semi-positivity constraint is met. In order to satisfy the total mass constraints, set

$$\tilde{p}_1^\epsilon = c_1^\epsilon p_1^* \exp\left[- \frac{1}{s_1^\epsilon} F_\epsilon'(\Lambda^\epsilon)\right] , \qquad (4.9)$$

$$\tilde{p}_0^\epsilon = c_0^\epsilon p_0^* \exp\left[- \frac{1}{s_0^\epsilon} G_\epsilon(\Lambda^\epsilon)\right] , \qquad (4.10)$$

where c_i^ϵ, $(i = 0, 1)$, are positive numbers chosen to satisfy the total mass constraints. Under C3, let $c_i^\epsilon \to c_i$ $(i = 0, 1)$ as $\epsilon \downarrow 0$.

Incorporating these constants, (4.6) becomes

$$\frac{1}{s_0^\epsilon} G_\epsilon(\Lambda^\epsilon) + \frac{1}{s_1^\epsilon} F_\epsilon'(\Lambda^\epsilon) + \ln \Lambda^\epsilon = \ln \Lambda^* + \ln \frac{c_1^\epsilon}{c_0^\epsilon}. \tag{4.11}$$

Via Lemma 3, we write

$$F_\epsilon'(\Lambda^\epsilon) = \frac{1}{T} G_\epsilon(\Lambda^\epsilon) + O(\epsilon) \tag{4.12}$$

and under condition C3 rewrite (4.11) as

$$\alpha^\epsilon G_\epsilon(\Lambda^\epsilon) + \ln \Lambda^\epsilon + O(\epsilon) = \ln \Lambda^* + \ln \frac{c_1}{c_0} \tag{4.13}$$

where $\alpha = ((s_1^\epsilon\ T)^{-1} + (s_0^\epsilon)^{-1})$. Let $\alpha^\epsilon \to \alpha$ as $\epsilon \downarrow 0$.

We now solve for $(\tilde{p}_0, \tilde{p}_1)$ for the three distinct cases: $\Lambda^\epsilon \in \Omega_0, \Omega_1, \Omega_2$ where $\Omega_0, \Omega_1, \Omega_2$ are defined by (3.20).

Case 1: $\Lambda^\epsilon \in \Omega_0$

For this case $G_\epsilon(\Lambda^\epsilon) = 0$ and (4.13) becomes

$$\ln \Lambda^\epsilon + O(\epsilon) = \ln \Lambda^* + \ln \frac{c_1^\epsilon}{c_0^\epsilon}. \tag{4.14}$$

As $\epsilon \downarrow 0$, then $\Lambda^\epsilon \to (c_1/c_0)\ \Lambda^*$ and (4.9)-(4.10) become

$$\tilde{p}_1 = c_1\, p_1^* \tag{4.15}$$

$$\tilde{p}_0 = c_0\, p_0^* \tag{4.16}$$

for $\tilde{p}_1/\tilde{p}_0 < T$ or equivalently $c_1\, p_1^*/(c_0\, p_0^*) < T$.

Case 2: $\Lambda^\epsilon \in \Omega_2$

For $\Lambda^\epsilon \in \Omega_2$, $G_\epsilon(\Lambda^\epsilon) = 1$ and (4.13) becomes

$$\alpha^\epsilon + \ln \Lambda^\epsilon + O(\epsilon) = \ln \Lambda^* + \ln \frac{c_1^\epsilon}{c_0^\epsilon}. \tag{4.17}$$

As $\epsilon \downarrow 0$, $\Lambda^\epsilon \to e^\alpha (c_1/c_0)\, \Lambda^*$ and (4.9)-(4.10) becomes

$$\tilde{p}_1 = c_1 e^{-\frac{1}{s_1 T}} p_1^* \tag{4.18}$$

$$\tilde{p}_0 = c_0 e^{\frac{1}{s_0}} p_0^* \tag{4.19}$$

for $\tilde{p}_1/\tilde{p}_0 \geq T$ or equivalently $c_1\, p_1^*/(c_0\, p_0^*) \geq Te^\alpha$.

Case 3: $\Lambda^\epsilon \in \Omega_1$

For this case (4.13) becomes

$$\frac{\alpha^\epsilon}{\epsilon} (\Lambda^\epsilon - T) + \ln \Lambda^\epsilon + O(\epsilon) = \ln \Lambda^* + \ln \frac{c_1^\epsilon}{c_0^\epsilon}. \tag{4.20}$$

Set $\Lambda^\epsilon = \Delta\Lambda + T$ where $\Delta\Lambda \geq 0$. Rewrite (4.20) as

$$\ln \left(1 + \frac{\Delta\Lambda}{T}\right) + \frac{\alpha^\epsilon}{\epsilon} \Delta\Lambda = \ln \frac{\Lambda^*}{T} + \ln \frac{c_1^\epsilon}{c_0^\epsilon} + O(\epsilon). \tag{4.21}$$

Now

$$\ln \left(1 + \frac{\Delta\Lambda}{T}\right) = \frac{\Delta\Lambda}{T} + O(\Delta\Lambda^2). \tag{4.22}$$

Using (4.21) and (4.22), it can be shown that

$$\Delta\Lambda = \frac{1}{\frac{1}{T} + \frac{\alpha}{\epsilon}} \left[\ln \frac{\Lambda^*}{T} + \ln \frac{c_1}{c_0} + O(\epsilon) + O(\Delta\Lambda^2) \right]. \tag{4.23}$$

Because $G_\epsilon(\Lambda^\epsilon) = \Delta\Lambda/\epsilon$, it follows from (4.23) that

$$G_\epsilon(\Lambda^\epsilon) = \frac{1}{\frac{\epsilon}{T} + \alpha}\left[\ln\frac{\Lambda^*}{T} + \ln\frac{c_1}{c_0} + O(\epsilon) + O(\Delta\Lambda^2)\right]. \tag{4.24}$$

As $\epsilon \downarrow 0$, because $\Delta\Lambda \le \epsilon$ then $O(\Delta\Lambda^2) \to 0$ and

$$G_\epsilon(\Lambda^\epsilon) \to \frac{1}{\alpha}\ln\left[\frac{c_1}{c_0 T}\Lambda^*\right]. \tag{4.25}$$

Using (4.12) and (4.25) in (4.9) and (4.10), respectively, it can be shown that

$$\tilde{p}_1 = c_1^\lambda c_0^{1-\lambda} T^{1-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} \tag{4.26}$$

$$\tilde{p}_0 = c_1^\lambda c_0^{1-\lambda} T^{-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} \tag{4.27}$$

where $\lambda \equiv (s_0\alpha)^{-1}$ and $T \le c_1 p_1^*/(c_0 p_0^*) < Te^\alpha$. Let $\lambda^\epsilon \equiv (s_0^\epsilon\alpha^\epsilon)^{-1}$ and $\lambda^\epsilon \to \lambda$ as $\epsilon \downarrow 0$ under condition C3.

Hence, we see that $(\tilde{p}_0^\epsilon, \tilde{p}_1^\epsilon) \to (\tilde{p}_0, \tilde{p}_1)$ pointwise (or in this case, uniformly on the finite space, Ω_x) as $\epsilon \downarrow 0$ under the conditions that 1) a solution exists for the unknown parameters of the densities, $(\tilde{p}_0, \tilde{p}_1)$, which are found via the constraint equations, 2) $(\tilde{p}_0, \tilde{p}_1) \in \text{int}\,(p^+)$ where p^+ is defined by (4.2b), and 3) condition C3 holds.

C. Conditions on the Solution

Now that we have the solution form for $\tilde{p}_0$ and $\tilde{p}_1$, we can give readily checked conditions under which condition C3 can be verified.

Lemma 4: C3 is true (i.e. all the parameters of the constraint equations have limits as $\epsilon \downarrow 0$) if

C4. a) $c_1 p_1^*/(c_0 p_0^*) \neq T$ or Te^α for all $\mathbf{x} \in \Omega_x$ and b) the Jacobian of the constraint equations is non-zero.

Proof: Condition C4 exemplifies the conditions associated with the Inverse Function Theorem [20]: the constraint equations must be continuously differentiable and the Jacobian of the constraint equations must be non-zero in order that the constraint equations are invertible in the neighborhood of c_0, c_1, α, and λ. It C4a is true, it is straightforward to show that the constraint equations are continuously differentiable. Set $y = (c_0^\epsilon, c_1^\epsilon, \alpha^\epsilon, \lambda^\epsilon)$ and $y_0 = (c_0, c_1, \alpha, \lambda)$ and let $z_i = f_i(y)$, $(i = 1, 2, 3, 4)$ denote the four constraint equations with $f_i(y_0) = 0$. Using the previous development of the derivation of $\tilde{p}_0$ and $\tilde{p}_1$, it can be shown that

$$\tilde{p}_1^\epsilon = \begin{cases} c_1^\epsilon p_1^* + O(\epsilon) & ; \; c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) < T + O(\epsilon) \\ c_1^{\lambda^\epsilon} c_0^{1-\lambda^\epsilon} T^{1-\lambda^\epsilon} p_1^{*\lambda^\epsilon} p_0^{*(1-\lambda^\epsilon)} + O(\epsilon) & ; \; T + O(\epsilon) \le c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) < Te^{\alpha^\epsilon} + O(\epsilon) \\ c_1^\epsilon e^{-(1-\lambda^\epsilon)\alpha^\epsilon} p_1^* + O(\epsilon) & ; \; c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) \ge Te^{\alpha^\epsilon} + O(\epsilon) \end{cases} \quad (4.28)$$

$$\tilde{p}_0^\epsilon = \begin{cases} c_0^\epsilon p_0^* + O(\epsilon) & ; \; c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) < T + O(\epsilon) \\ c_1^{\lambda^\epsilon} c_0^{1-\lambda^\epsilon} T^{-\lambda^\epsilon} p_1^{*\lambda^\epsilon} p_0^{*(1-\lambda^\epsilon)} + O(\epsilon) & ; \; T + O(\epsilon) \le c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) < Te^{\alpha^\epsilon} + O(\epsilon) \\ c_0^\epsilon e^{\lambda^\epsilon \alpha^\epsilon} p_0^* + O(\epsilon) & ; \; c_1^\epsilon p_1^*/(c_0^\epsilon p_0^*) \ge Te^{\alpha^\epsilon} + O(\epsilon) \end{cases} \quad (4.29)$$

where the order terms added to each T and Te^{α^ϵ}, respectively are identical. Examining (4.28) and (4.29), it is found that expressions for $\tilde{p}_i^\epsilon (i = 1, 2)$ at the boundaries of the regions of applicability are within $O(\epsilon)$. Thus (4.28) and (4.29) can be rewritten such that the order terms do not appear in the regions of applicability but are incorporated as order terms in the expressions for $\tilde{p}_i^\epsilon (i = 0, 1)$.

If these solutions are substituted into the constraint equations, it is found that $f_i(y) = O_i(\epsilon)$ $(i = 1, 2, 3, 4)$ where we have subscripted each ordered term to indicate its distinctness. We point out that each constraint equation has three summations, each taken over one of the three regions defined by (4.28) or (4.29). Condition C4a guarantees that a term appearing in one of the summations will not jump to another summation for arbitrarily small perturbations about y_0. This will also be true for the first and second derivitives of $f_i (i = 1, 2, 3, 4)$. Under C4a, it can be shown that each term of each summation of $f_i (i = 1, 2, 3, 4)$ is continuously differentiable and hence f_i is continuously differentiable. Hence under C4 and the Inverse Function Theorem, the solution for $(c_0^\epsilon, c_1^\epsilon, \alpha^\epsilon, \lambda^\epsilon)$ exists in the neighborhood of $c_0, c_1, \alpha, \lambda$ for arbitrarily small ϵ and $\lim_{\epsilon \downarrow 0} (c_0^\epsilon, c_1^\epsilon, \alpha^\epsilon, \lambda^\epsilon) = (c_0, c_1, \alpha, \lambda)$.□

We point out that it is highly likely that C4 is true.

The condition that $\infty > s_0, s_1 \ge 0$ is equivalent to the condition $\alpha > 0$ and $0 < \lambda < 1$. This can be shown via the equations: $\alpha = (s^1 T)^{-1} + s_0^{-1}$ and $\lambda = (s_0 \alpha)^{-1}$. Because $p_0^* > 0$ and Ω_x is compact, it follows that condition C2 of Theorem 2 holds. Under condition C4 and the preceding development we see that condition C1 of Theorem 2 holds. Thus we can state:

Theorem 4: The least favorable T-dependent pair for the divergence class discrimination hypothesis testing problem is given by the following densities under the conditions 1) that a solution exists for the unknown parameters of the densities which are found via the constraint equations (and the total mass constraints), 2) $(\tilde{p}_0, \tilde{p}_1) \in \text{int}\,(p^+)$ where p^+ is defined by (4.2b) 3) C4 holds and 4) $\alpha > 0$, $0 < \lambda < 1$:

$$\tilde{p}_1 = \begin{cases} c_1\, p_1^* & ;\ c_1\, p_1^*/(c_0\, p_0^*) < T \\ c_1^{\lambda} c_0^{1-\lambda} T^{1-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} & ;\ T \le c_1\, p_1^*/(c_0\, p_0^*) < Te^{\alpha} \\ c_1 e^{-(1-\lambda)\alpha} p_1^* & ;\ c_1\, p_1^*/(c_0\, p_0^*) \ge Te^{\alpha} \end{cases}$$

$$\tilde{p}_0 = \begin{cases} c_0\, p_0^* & ;\ c_1\, p_1^*/(c_0\, p_0^*) < T \\ c_1^{\lambda} c_0^{1-\lambda} T^{-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} & ;\ T \le c_1\, p_1^*/(c_0\, p_0^*) < Te^{\alpha} \\ c_0 e^{\lambda\alpha} p_0^* & ;\ c_1\, p_1^*/(c_0\, p_0^*) \ge Te^{\alpha} \end{cases}$$

where c_0, c_1, λ, α are the unknown parameters to be determined from the four constraint equations. The likelihood ratio $\tilde{p}_1/\tilde{p}_0$ is given by

$$\tilde{\Lambda} = \begin{cases} c_1\, p_1^*/(c_0\, p_0^*) & ;\ c_1\, p_1^*/(c_0\, p_0^*) < T \\ T & ;\ T \le c_1\, p_1^*/(c_0\, p_0^*) < Te^{\alpha} \\ e^{-\alpha} c_1\, p_1^*/(c_0\, p_0^*) & ;\ c_1\, p_1^*/(c_0\, p_0^*) \ge Te^{\alpha}. \end{cases} \tag{4.30}$$

The decision rule is given by

$$\tilde{\phi} = \begin{cases} 0 & ;\ p_1^*/p_0^* \le c_0 Te^{\alpha}/c_1 \\ 1 & ;\ p_1^*/p_0^* > c_0 Te^{\alpha}/c_1. \end{cases} \tag{4.31}$$

D. Calculation of P_D, P_F

Let $P_D^*(T)$ and $P_F^*(T)$ be the probabilities of detection and false alarm for $P_1^* \sim H_1$ and $P_0^* \sim H_0$, respectively of the nominal decision rule with threshold, T. Using (4.31)

$$P_D(\tilde{P}_1, \tilde{\phi}, T) = \text{Prob}_{\tilde{p}_1} \left\{ \frac{c_1}{c_0} e^{-\alpha} \frac{p_1^*}{p_0^*} > T \right\}$$

$$= \sum_{\left(\frac{c_1}{c_0} e^{-\alpha} \frac{p_1^*}{p_0^*} > T\right)} c_1 \, e^{-(1-\lambda)\alpha} \, p_1^* \; d\mu$$

$$= c_1 \, e^{-(1-\lambda)\alpha} \, P_D^* \left[\frac{c_0 T e^{\alpha}}{c_1} \right]. \tag{4.32}$$

Similarly, it can be shown

$$P_F(\tilde{P}_0, \tilde{\phi}, T) = c_0 \, e^{\lambda\alpha} \, P_F^* \left[\frac{c_0 T e^{\alpha}}{c_1} \right]. \tag{4.33}$$

Thus for any $(P_0, P_1) \in \mathcal{P}$, using (3.53), (4.30), and (4.31)

$$R_B(P_0, \tilde{\phi}, T) \le \pi_0 \, C_0 c_0 e^{\lambda\alpha} \, P_F^* \left[\frac{c_0 T e^{\alpha}}{c_1} \right]$$

$$+ \pi_1 C_1 \left[1 - c_1 \, e^{-(1-\lambda)\alpha} \, P_D^* \left[\frac{c_0 T e^{\alpha}}{c_1} \right] \right]. \tag{4.34}$$

Hence knowing T, c_0, c_1, λ, $P_D^*(\cdot)$, and $P_F^*(\cdot)$ allows us to find the upper-bound on Bayes risk over the uncertainty classes of the two hypotheses.

V. EXAMPLE: DIVERGENCE/LINEAR CLASS

The results of the previous section can be readily extended to a class we call the Divergence/Linear class (or simply D/L class). In general, the two hypothesis classes are defined as

$$\mathcal{P}_i = \{ P_i \mid D(p_i, p_i^*) \le \Delta_i ; \int_X h_{im}(\mathrm{x}) p_i(\mathrm{x}) d\mu = c_{im} ; \; m = 1,2,\ldots M \} ; \; i = 0,1 \; , \tag{5.1}$$

where h_{im} are known functions of the elements of x and the c_{im} are specified. For example, if $h_{im}(x)$ is a multidimensional monomial of the elements of x then c_{im} corresponds to a moment of these elements. We again assume $\mathcal{P}_0 \cap \mathcal{P}_1 = \Phi$. In addition, we assume that the constraints indicated by (5.1) are regular [17] on Ω_x. Because the constraints are convex functions of p_i (i = 0, 1) it is straightforward to show that $\mathcal{P}_i$ (i = 0, 1) are convex sets. It is shown in Appendix D that the D/L class is not necessarily 2-alternating capacitable.

As in the preceding sections, we restrict Ω_x to have a finite number of elements. However in this case, due to the moment constraints, a support must be specified. Set $\Omega_x = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K\}$ where there are K elements in the support and $\mathbf{x}_k$ ($k = 1, 2, \ldots, K$) is an N-length vector and $K \geq M + 3$. Define

$$\mathcal{p}_i = \{p_i \mid D(p_i, p_i^*) \leq \Delta_i;\ \sum_I h_{im}\, p_i = c_{im};\ m = 1,2,\ldots,M;\ p_i \in \mathcal{P}^K\},\ i = 0,1 \quad (5.2a)$$

$$\mathcal{p}_i^+ = \{p_i \mid D(p_i, p_i^*) = \Delta_i;\ \sum_I h_{im}\, p_i = c_{im};\ m = 1,2,\ldots,M;\ p_i \in \mathcal{P}^K\},\ i = 0,1. \quad (5.2b)$$

We assume $\boldsymbol{p}_i^+$ contains an infinite number of elements which will be true if $K \geq M + 3$. Again (as in the previous section), $\boldsymbol{p}$ and $\boldsymbol{p}^+$ are compact and Theorem 1 guarantees that the minimum to (3.4) on $\boldsymbol{p}$ or $\boldsymbol{p}^+$ exists. As before we assume the least favorable T-dependent densities is on the interior of $\boldsymbol{p}^+$ or $(\tilde{p}_1, \tilde{p}_2) \in \operatorname{int}(\boldsymbol{p}^+)$.

The Lagrangian for the minimization problem posed by (3.3) and the class given by (5.2a) (or 5.2b) is

$$L = \sum_I \tilde{p}_0^\epsilon F\left[\frac{\tilde{p}_1^\epsilon}{\tilde{p}_0^\epsilon}\right] + s_0^\epsilon \sum_I \tilde{p}_0^\epsilon \ln \frac{\tilde{p}_0^\epsilon}{p_0^*} + s_1^\epsilon \sum_I \tilde{p}_1^\epsilon \ln \frac{\tilde{p}_1^\epsilon}{p_1^*}$$
$$+ \sum_I \sum_{m=1}^{M} (\lambda_{0m} h_{0m}\, \tilde{p}_0^\epsilon + \lambda_{1m} h_{1m}\, \tilde{p}_1^\epsilon) \quad (5.3)$$

where $s_i\ \lambda_{im}$ are the Lagrange multipliers ($i = 0, 1$; $m = 1, 2, \ldots, M$) and each $\tilde{p}_0^\epsilon$, $\tilde{p}_1^\epsilon$, p_0^*, p_1^*, h_{0m}, h_{1m} are indexed with respect to the elements of I.

Using the methodology of the previous section it is straightforward to show:

Theorem 5: The least favorable T-dependent pair for the divergence/linear class discrimination problem is given by the following densities under the conditions that 1) a solution exists for the unknown parameters of the densities which are found via the constraint equations ((4.5), total mass constraints, and moment constraints), 2) $(\tilde{p}_0, \tilde{p}_1) \in \operatorname{int}(\boldsymbol{p}^+)$ where $\boldsymbol{p}^+$ is defined by (5.2b), 3) C5 holds (given below) and 4) $\alpha > 0$, $0 < \lambda < 1$:

Define V (which is indexed by I) as

$$V = \frac{c_1\, p_1^*}{c_0\, p_0^*} \exp\left\{ - \sum_{m=1}^{M} (a_{1m} h_{1m} - a_{0m} h_{0m}) \right\}$$

and the sets

$$\Omega_{V_0} = \{k | V < T\}$$

$$\Omega_{V_1} = \{k | T \le V < Te^{\alpha}\}$$

$$\Omega_{V_2} = \{k | V \ge Te^{\alpha}\}$$

$$\Omega_{V_2}^{+} = \{k | V > Te^{\alpha}\}.$$

Condition C5 (which is an extension of Lemma 4) is

C5. a) $V \neq T$ or Te^{α} for all x $\in \Omega_x$ and

b) the Jacobian of the constraint equations is non-zero.

The densities are given by

$$\tilde{p}_1 = \begin{cases} c_1 p_1^* \exp\left\{ -\sum_{m=1}^{M} a_{1m} h_{1m} \right\} & ; \; k \in \Omega_{V_0} \\ c_1^{\lambda} c_0^{1-\lambda} T^{1-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} \exp\left\{ -\sum_{m=1}^{M} [(1-\lambda) a_{0m} h_{0m} + \lambda a_{1m} h_{1m}] \right\} & ; \; k \in \Omega_{V_1} \\ c_1 e^{-(1-\lambda)\alpha} p_1^* \exp\left\{ -\sum_{m=1}^{M} a_{1m} h_{1m} \right\} & ; \; k \in \Omega_{V_2} \end{cases}$$

$$\tilde{p}_0 = \begin{cases} c_0 p_0^* \exp\left\{ -\sum_{m=1}^{M} a_{0m} h_{0m} \right\} & ; \; k \in \Omega_{V_0} \\ c_0^{\lambda} c_0^{1-\lambda} T^{-\lambda} p_1^{*\lambda} p_0^{*(1-\lambda)} \exp\left\{ -\sum_{m=1}^{M} [(1-\lambda) a_{0m} h_{0m} + \lambda a_{1m} h_{1m}] \right\} & ; \; k \in \Omega_{V_1} \\ c_0 e^{\lambda\alpha} p_0^* \exp\left\{ -\sum_{m=1}^{M} a_{0m} h_{0m} \right\} & ; \; k \in \Omega_{V_2} \end{cases}$$

where c_i, λ, α, a_{im}, ($i = 0, 1$; $m = 1, 2, \ldots, M$) are unknown parameters to be determined.

The likelihood ratio of $\tilde{p}_1/\tilde{p}_0$ is given by

$$\Lambda = \begin{cases} V & ; \; k \in \Omega_{V_0} \\ T & ; \; k \in \Omega_{V_1} \\ e^{-\alpha}V & ; \; k \in \Omega_{V_2} \end{cases} \tag{5.4}$$

The decision rule is given by

$$\tilde{\phi} = \begin{cases} 0 & ; \; k \in \Omega_{V_2}^{+c} \\ 1 & ; \; k \in \Omega_{V_2}^{+} \end{cases} \tag{5.5}$$

where c denotes the complement set. We note that Theorem 4 is actually a special case of Theorem 5 because the total mass constraints can be written as linear constraints with h_i, $= 1$, $(i = 0, 1)$.

VI. SUMMARY

We have presented a methodology for finding robust detectors for composite binary hypotheses defined for uncertainty classes which are not necessarily 2-alternating capacitable. A robust detector is defined as a detection structure whose performance measures are sharply lower and/or upperbounded for given input uncertainty classes. Past robust detection schemes have been threshold independent. The robust test reduced to a test between two simple hypotheses whereby the underlying probability measures were fixed representatives of the specified uncertainty classes and were independent of the test's threshold. In this paper, we presented conditions and formulations for detection structures which can be threshold dependent, and which sharply upper-bound the Bayes risk for the chosen detector threshold. The support set was assumed to have a finite number of elements. The robust detector structure resulted from solving an associated limiting minimization problem. It was shown that the robust test again reduces to a test between two simple hypotheses whereby the underlying probability measures were fixed representatives of the specified uncertainty classes. However. these probability measures can be a function of the detector's threshold. Results on the existence of these solutions were presented and solutions for the divergence and divergence/linear uncertainty classes were formulated.

REFERENCES

1. V. Strassen, Messfehler und information, *Z. Wahrsch. verw. Gebeite*, **2** 273-305, 1964.

2. V. Strassen, The existence of probability measures with given marginals, *Ann. Math. Statist.* **36**, 423-439, 1965.

3. P.J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capabilities," *Ann. Stat.*, vol. 1, pp. 251-263, 1973, and *Ann. Stat.*, vol. 2, pp. 223-224, 1974.

4. G. Choquet, Theory of capacities, *Ann. Inst. Fourier* **5**, 131-292, 1953-54.

5. G. Choquet, Forme abstraite du théoreme de capacitabilité, *Ann. Inst. Fourier* **9**, 83-89, 1959.

6. H. Rieder, "Least favorable pairs for special capacities," *Ann. Statist.* **5**, 909-921, 1977.

7. T. Bednarski, "On Solutions of Minimax Test Problems for Special Capacities," *Z. Wahrscheinlichkeitheorie verw. Gebiete.* **46**, pp. 397-405, 1981.

8. K.S. Vastola, "Topics in Robust Signal Processing," Phd. dissertation, University of Illinois, 1982.

9. P.J. Huber, "A robust version of the probability ratio test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.

10. S.A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. on Inform. Theory*, vol. IT-27, No. 2, pp. 242-247, 1981.

11. K.S. Vastola and H.V. Poor, "On the p-point uncertainty class," *IEEE Trans. on Inform. Theory*, vol. IT-30, pp. 374-376, 1984.

12. H.V. Poor, "Robust decision design using a distant criteria," *IEEE Trans. on Inform. Theory*, vol. IT-26, No. 5, pp. 575-587, 1980.

13. D.S. Sauder and E. Geraniotis, "Optimal and robust memoryless discrimination from dependent observations," *IEEE Trans. on Inform. Theory*, vol. IT-37, no. 1, pp. 73-91, 1991.

14. H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1988.

15. P.J. Huber, *Robust Statistics*, J. W. Wiley and Sons, New York, 1981.

16. V. Barbu and Th. Precupanu, *Convexity and Optimization in Banach Spaces*, Bucarest: Editura Academiei, 1978.

17. R. Pallu de la Barriere, *Optimal Control Theory*, Dover, New York, 1967.

18. M. Minoux, *Mathematical Programming*, J. W. Wiley and Sons, New York, 1986.

19. S. Kullback, *Information Theory and Statistics*, J. W. Wiley and Sons, New York, 1959.

20. W. Rudin, *Principles Mathematical Analysis*, p. 196, McGraw Hill, New York, 1964.

21. R.E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. on Inform. Theory*, vol. IT-20, No. 4, pp. 405-417, 1974.

Appendix A
PROOF OF LEMMA 2

The regions Ω_0, Ω_1, and Ω_2 are defined by (3.20). If $z_{max} < T + \epsilon$, the following arguments are easily modified to obtain the proof. Thus we assume $z_{max} \geq T + \epsilon$. The functions. $F_\epsilon(z, T)$ and $F_0(z, T)$ are defined by (3.21) and (3.22), respectively. Define

$$D_\epsilon(z, T) = F_\epsilon(z, T) - F_0(z, T) . \tag{A1}$$

To demonstrate that $F_\epsilon(z, T) \rightarrow F_0(z, T)$ uniformly on $[0, z_{max}]$ as $\epsilon \downarrow 0$, we show that $\sup_z \mid D_\epsilon(z, T) \mid \rightarrow 0$ as $\epsilon \downarrow 0$ in each of the regions: $0 \leq z \leq T$ and $T < z \leq z_{max}$ ($|\cdot|$ denotes absolute value).

A. For $z \leq T$, $D_\epsilon(z, T) = 0$ and it trivially follows that $\sup_z \mid D_z(z, T) \mid = 0$.

B. For $z > T$, set $z = T + \Delta$ and choose $\epsilon < \Delta$. For $z \in [T + \epsilon, z_{max}]$

$$D_\epsilon(z, T) = \frac{z}{\epsilon}\left[\ln\left(1 + \frac{\epsilon}{T}\right) - \frac{\epsilon}{T}\right] . \tag{A2}$$

Since $D_\epsilon(z, T)$ is linear with a negative slope and zero intercept, the maximum occurs at z_{max}. It is straightforward to show $\mid D_\epsilon(z_{max}, T \mid = O(\epsilon)$. Hence $\sup_z \mid D_\epsilon(z, T) \mid \rightarrow 0$ as $\epsilon \downarrow 0$ for $z > T$.

Appendix B
PROOF OF LEMMA 3

The regions Ω_0, Ω_1, Ω_2 and the function $G_\epsilon(z, T)$ are defined in Eq. (3.20). Now

$$F_\epsilon' = \int_0^z \frac{G_\epsilon(\beta, T)}{\beta^2} \, d\beta + \frac{G_\epsilon(z, T)}{z} . \tag{B1}$$

Define

$$H_\epsilon(z, T) = F_\epsilon'(z, T) - \frac{1}{T} G_\epsilon(z, T) . \tag{B2}$$

To demonstrate that $H_\epsilon(z, T)$ converges uniformly to 0 for $z \geq 0$ we show that $\sup_z | H_\epsilon(z, T) | \to 0$ as $\epsilon \downarrow 0$ in each region $z < T$, $z = T$, $z > T$.

A. For $z < T$, $H_\epsilon(z, T) = 0$ and it trivially follows that $\sup_z | H_\epsilon(z, T) | = 0$ in this region.

B. For $z \in \Omega_1$ we can show

$$H_\epsilon(z, T) = \frac{1}{\epsilon} \ln \frac{z}{T} - \frac{1}{\epsilon T}(z - T) . \tag{B3}$$

Thus $H_\epsilon(T, T) = 0$.

C. For $z > T$, we set $z = T + \Delta$ and choose $\epsilon < \Delta$. Thus $z \in \Omega_2$ and it can be shown

$$F_\epsilon'(z, T) = \frac{1}{\epsilon} \ln \frac{T + \epsilon}{T} . \tag{B4}$$

Now $H_\epsilon(z, T)$ is independent of z and

$$F_\epsilon'(z, T) - \frac{1}{T} G_\epsilon(z, T) = \frac{1}{\epsilon} \ln \frac{T + \epsilon}{T} - \frac{1}{T} = O(\epsilon) . \tag{B5}$$

Thus $| H_\epsilon(z, T) |$ is uniformly convergent to 0 for $z > T$.

Appendix C
CONDITION FOR NON-OVERLAPPING DIVERGENCE CLASSES

Let $\Delta_1 = \Delta_0 = \Delta$. We wish to find the minimum Δ such that there exists a pdf p, satisfying

$$\sum_{\Omega_x} p \ln \frac{p}{p_i^*} \leq \Delta \ , \ i = 0, 1 \ . \tag{C1}$$

It is straightforward to show that this is equivalent to finding

$$\Delta_{\min} = \min_p \sum_{\Omega_x} p \ln \frac{p}{p_i^*} \qquad i = 0, 1 \tag{C2}$$

subject to the constraint

$$\sum_{\Omega_x} p \ln \frac{p}{p_0^*} = \sum_{\Omega_x} p \ln \frac{p}{p_1^*} \ . \tag{C3}$$

It can be shown that

$$p = \frac{{p_1^*}^{1-s} \, {p_0^*}^{s}}{\sum_{\Omega_x} {p_1^*}^{1-s} \, {p_0^*}^{s}} \tag{C4}$$

where s is determined from the constraint equation, Eq. (C3). Substituting and simplifying implies s is the solution of

$$\sum_{\Omega_x} {p_1^*}^{1-s} \, {p_0^*}^{s} \ln \frac{p_1^*}{p_0^*} = 0 \ . \tag{C5}$$

Set

$$f(s) = \sum_{\Omega_x} {p_1^*}^{1-s} \, {p_0^*}^{s} \ln \frac{p_1^*}{p_0^*} . \tag{C6}$$

We note $f(s)$ is continuous and monotonically decreasing, on the internal $s \in [0, 1]$. In addition, $f(0) > 0$ and $f(1) < 0$. Thus a solution exists and

$$\Delta_{\min} = -\ln \sum_{\Omega_x} {p_1^*}^{1-s} \, {p_0^*}^{s} . \tag{C7}$$

Hence for $\Delta < \Delta_{\min}$ the divergence uncertainty classes do not overlap.

Appendix D
THE DIVERGENCE OR DIVERGENCE/LINEAR CLASSES NEED NOT BE 2-ALTERNATING CAPACITABLE

We need only consider the D/L classes since the div-class is embedded in the D/L class. We can construct a D/L class that has only one member in it. Because this class is trivially 2-alternating capacitable, we cannot be definitive and say no D/L class is 2-alternating capacitable.

Assume that $\mathcal{P}_1$ is defined by Eq. (5.1) and that this class is 2-alternating capacitable. Specifically, let $\mathcal{P}_1$ be div-class with finite support such that $K \geq 2$. Let $\mathcal{P}_0$ have only one member, P_0, (and thus is 2-alternating capacitable) such that $P_0 \notin \mathcal{P}_1$. It is straightforward to modify Huber and Strassen's [3] Theorem 6.1 to show that if F is any twice continuously differentiable function on $(0, \infty)$ and $P_0 >> P_l$ then the least favorable pair (in the sense of Huber and Strassen), each of which is 2-alternating capacitable, minimizes $J(p_0, p_1)$ which defined by (3.3). The resultant pair is functionally independent of the choice of F. Consider the solutions for $\tilde{p}_1$ under $F = F_1 = -\ln z$ and $F = F_2 = z \ln z$. Both F_1 and F_2 are convex and twice continuously differentiable on $(0, \infty)$. For F_2, the solution exists and is given by Blahut [21]. Under F_1, it can be shown that a solution exists for some ϵ_1. However, the solutions under F_1 and F_2 are not identical. Hence $\mathcal{P}_1$ is not capacitable.